

Open Source.

Exploring opportunities to commercialize open source by leveraging data.

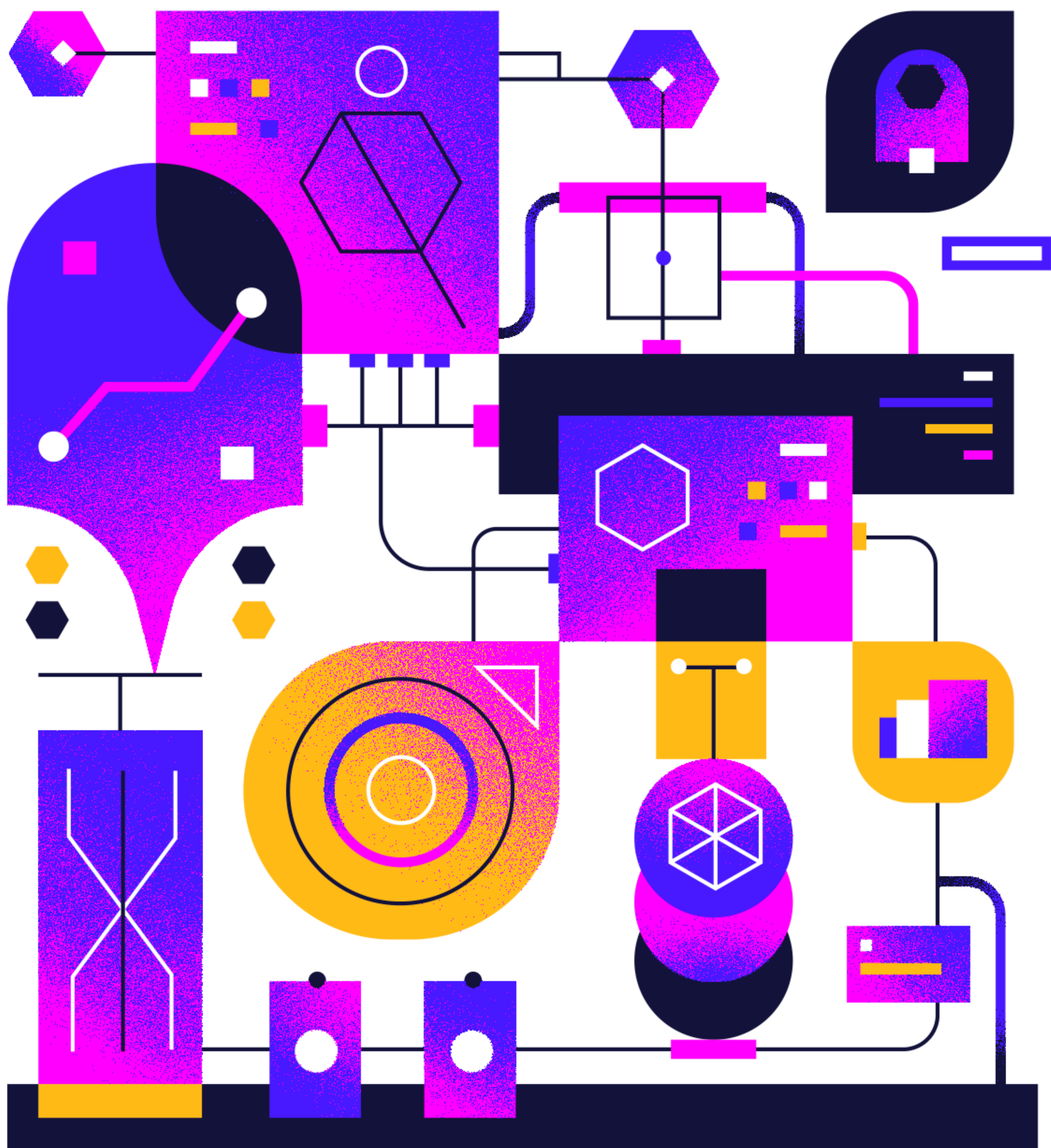
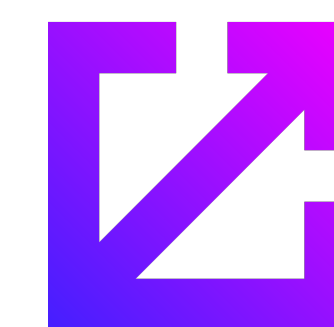


Table of contents.

Open Source: moving from exception to mainstream	4
I got 99 Issues and that's a good thing	5
Contributors and the Coverage Ration: The inside scoop	7
Three metrics to rule them all	8
Methodology	9



The monetization of Open Source...

...is laying the foundation for the next set of iconic tech companies. As open source has matured from an ideology into a go-to-market motion, category-defining companies are being built on open source, like MongoDB, Elastic and GitHub, to name a few.

Since late 2017 when we led the seed round for Rasa, an open source conversational AI platform, we've been tracking and working to understand the most important drivers behind open source successes. We analyzed data from the hottest new open source companies to help founders make sense of early metrics and determine what to prioritize. We are starting with data on GitHub, where almost all open source projects live, and will layer on data from other channels — e.g., downloads, installs, Hackernews, Reddit, Stackoverflow, Twitter — in future reports.

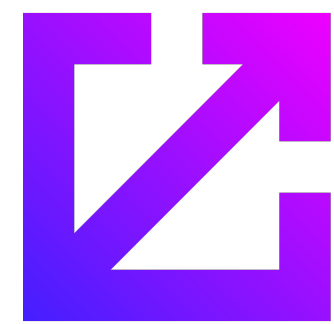
For founders building open source companies, they must continuously and accurately assess the awareness and engagement of external developers, who adopt your product as users and some may also become contributors. Adoption is universally important. Getting contributions from your community comes into play later on and may be less important depending on your product, sector and strategy. Nothing is more telling of a company's ability to win on these dimensions than Issues, Serious Contributors, or the Contributor Coverage Ratio.



Written by Chang Xu

Partner at





Open source: moving from exception to mainstream

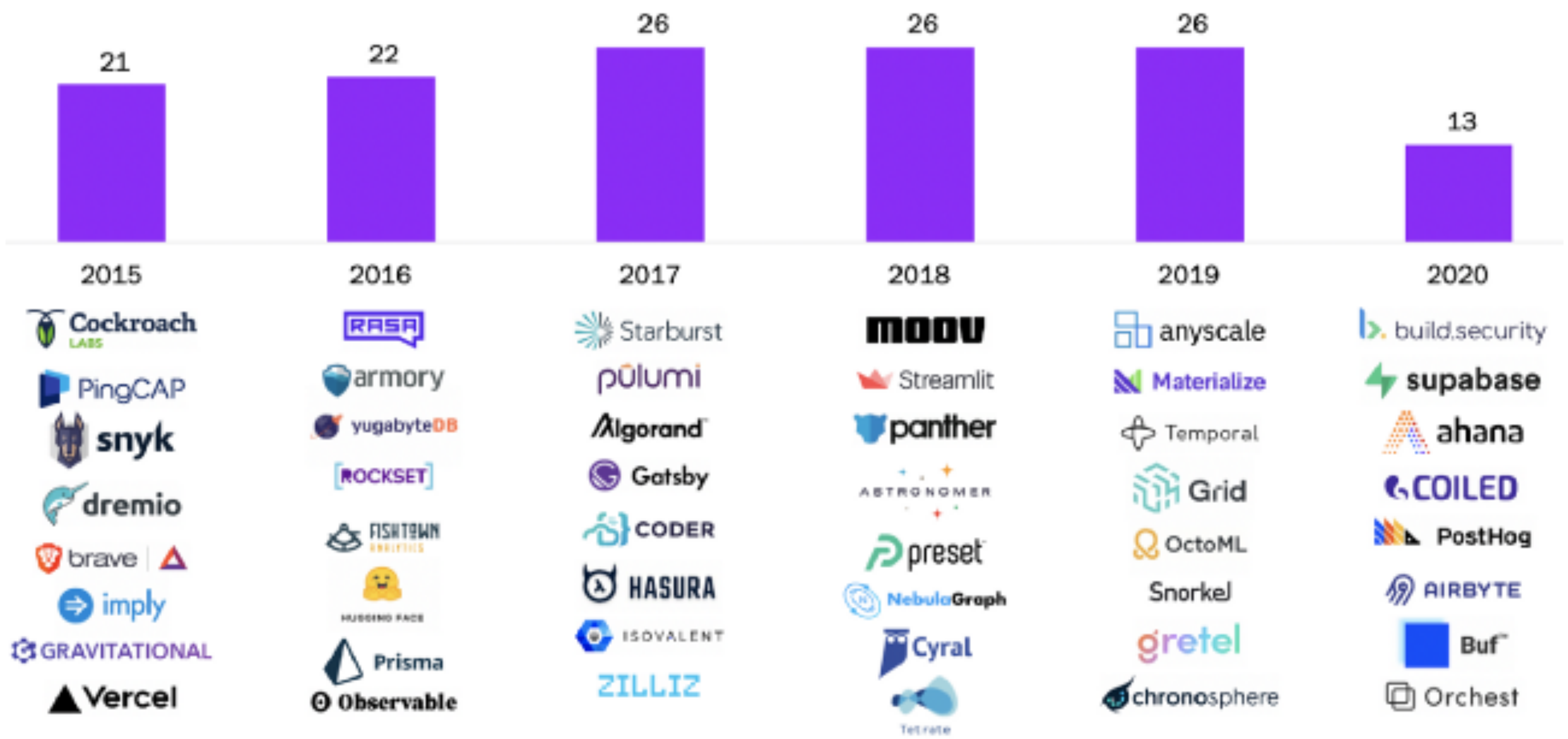
The Basis Set Open Source Index includes 134 companies founded since 2015 that are commercializing open source and have raised at least \$1M in venture funding. More open source companies are founded every year, from 21 in 2015 to 26 in 2019 [1].

More companies are also deliberately using open source as go-to-market strategy. In 2015, 67% of companies that launch open source launch it much later after company formation. By 2020, this statistic has drastically dropped to 15%. By contrast, the proportion of companies that were formed concurrently as their open source project have risen from 5% to 38% [2]. Companies are often reaching for open source right out of the gate.



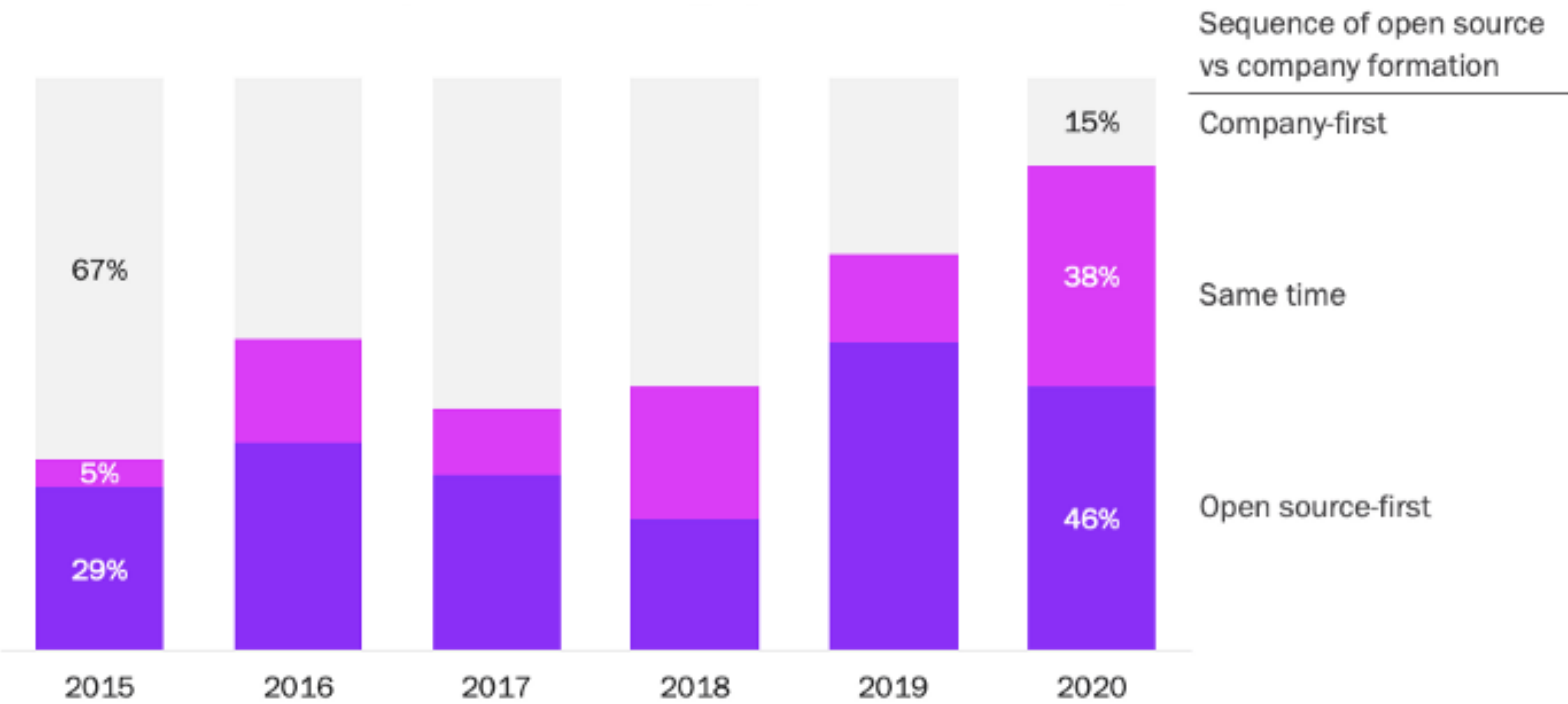
Basis Set Open Source Index

By company founding year





More companies are using open source from day one



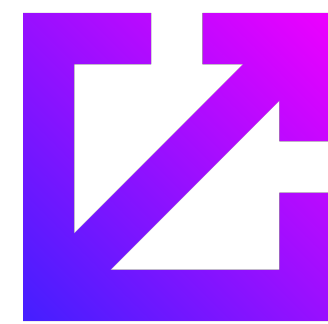
As a rough proxy for success, we used total funding raised [3]. We looked at how companies’ open source repos perform across different funding tiers from \$1–5M raised to \$50M+ raised, roughly correlating with seed stage to growth stage.

I got 99 Issues and that’s a good thing

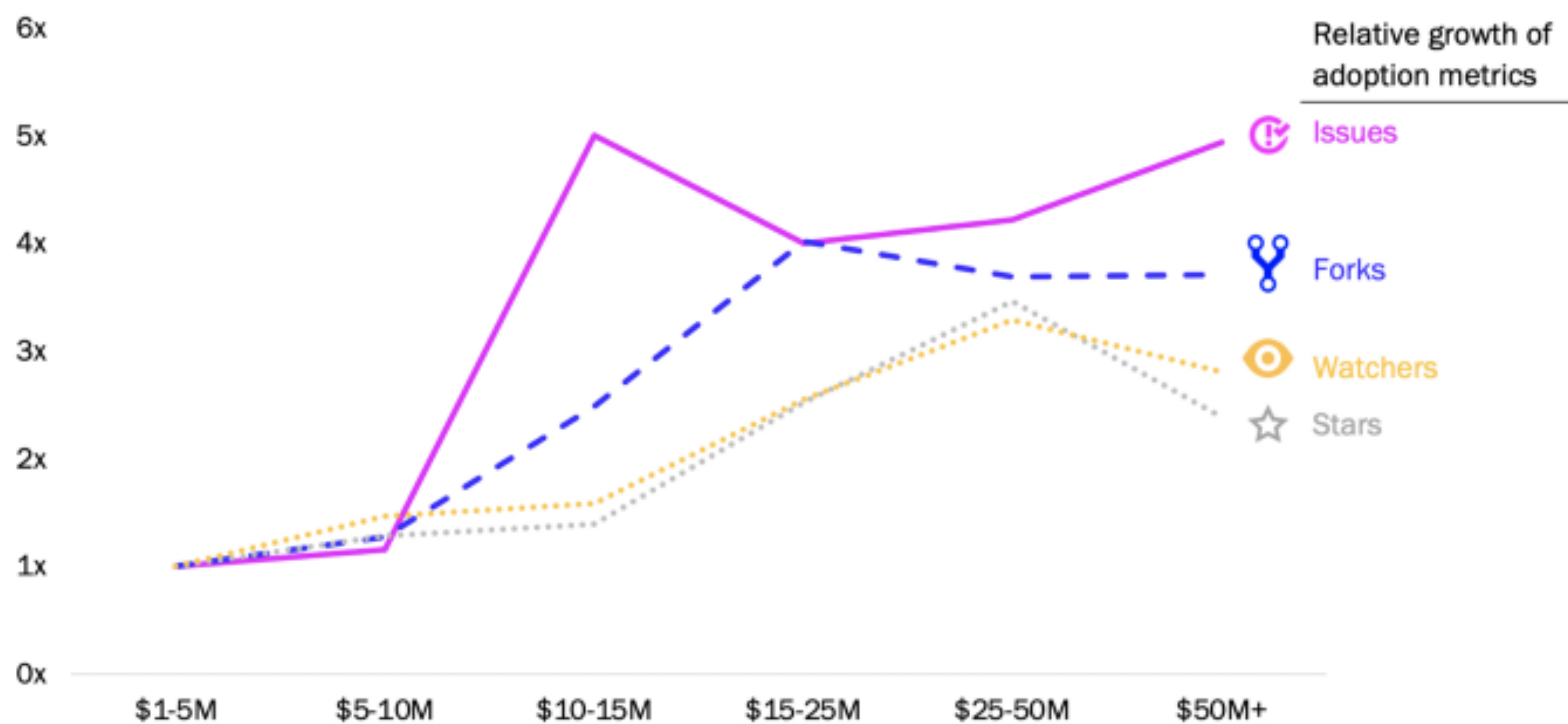
When looking at adoption, Issues is the name of the game. It grows the fastest and most consistently among all the metrics, because it shows that external developers are actually engaging with your product. Growth stage companies typically have 5x the number of Issues as seed stage companies.

Rick Lamers, the cofounder & CEO of Orchest, mentioned that when he was building the popular open source project Grid Studio, he noticed that the number of Issues jumped at around 6,000 stars. It was clear that he had active users and they were running into real issues.

As you build in public, you can naturally have your Issues flow into your product roadmap publicly, as GitHub and Cockroach do. Other best practices for engaging your community include writing high quality documentation, detailed release notes, like Spinnaker, or changelogs, like Rasa and Pulumi. Being open with your community fosters trust.



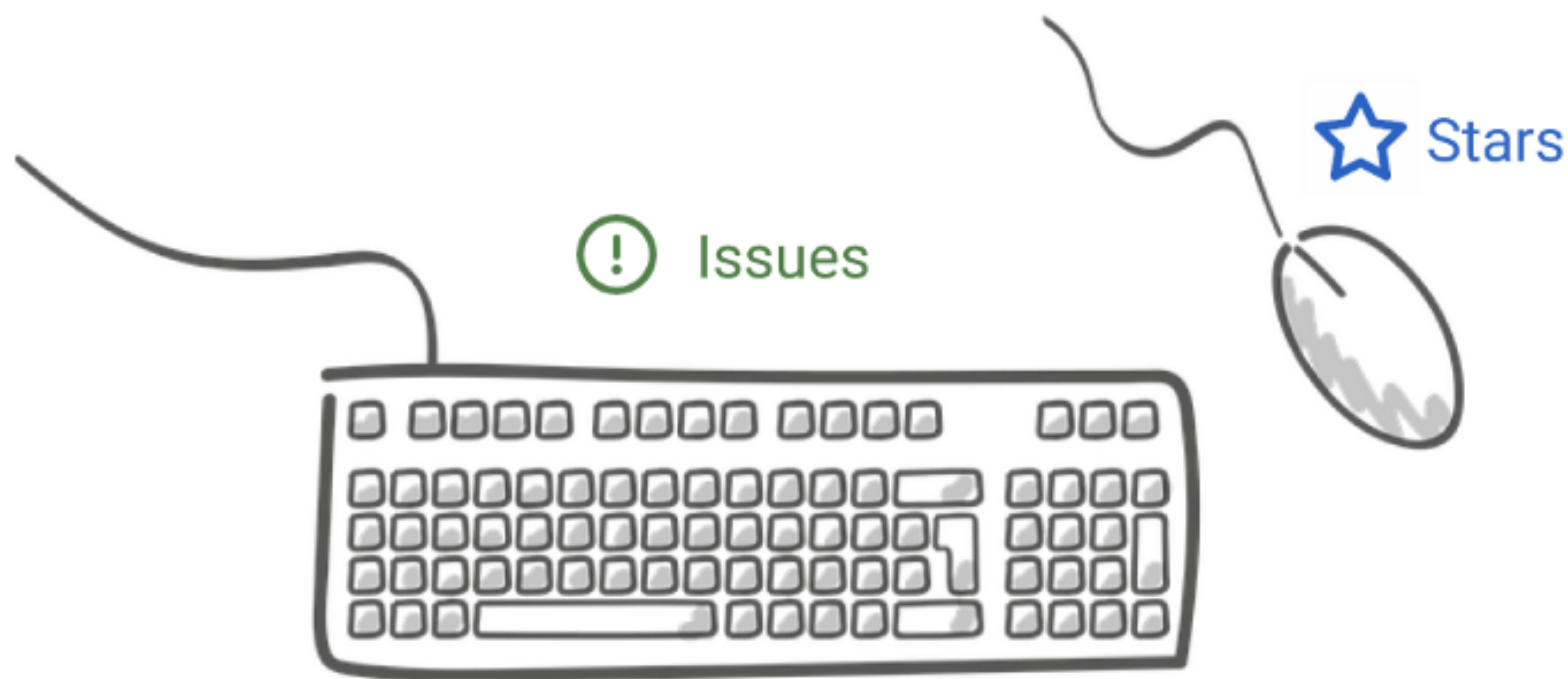
To understand adoption, issues is the most important metric



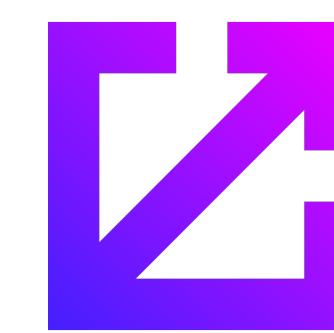
Forks grow the next fastest, at 4x, because it indicates a bit more active engagement than stars or watchers. It is a prerequisite for actually playing with the code or building something on top of it.

Stars and Watchers grow the most slowly and have more variability, at 2–3x. They're perhaps the least interesting as metrics because they track interest passively.

In conversation with Jason Warner, CTO of GitHub, he offered a fun analogy for understanding this: there are two inputs to a computer — a keyboard and a mouse. He values the keyboard and doesn't value the mouse. Typing up Issues indicates a far stronger interest than clicking to Star a repo.



The keyboardd (Issues) is more valuable than the mouse (Stars)

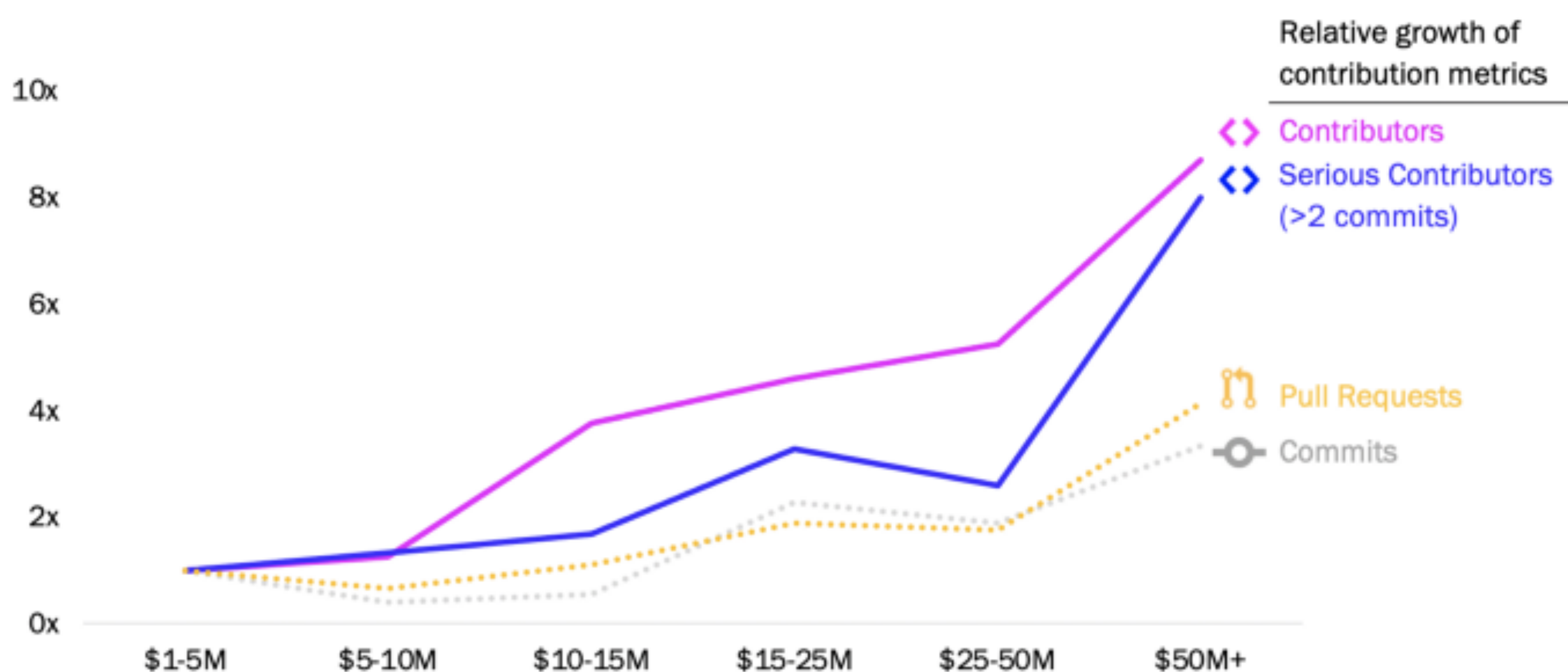


Contributors and the Coverage Ratio: the inside scoop

Moving from adoption to contribution often signals the next level of engagement, though some projects strategically choose to focus only on user growth and not on contributor growth, such as Babel [4]. In looking at metrics about contribution, we excluded the companies that were open source-first as their original open source communities — some very large and they often do not have a lot of control over — gave them very different starting points.

The number of Contributors grows at the fastest rate. In fact, the number of Serious Contributors (who have made >2 commits) rises sharply as companies get into the growth stage. Growth stage companies have 9x as many Contributors and 8x as many Serious Contributors as seed stage companies. Pull Requests and Commits grow at a more modest rate of 3–4x.

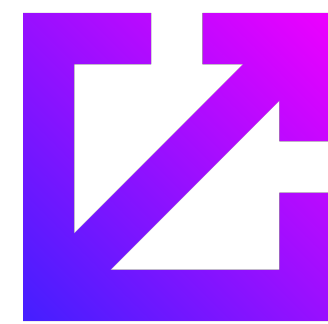
To understand contribution, Contributor and Serious Contributor count are the most important metrics



To understand why contributor growth is so important, we looked at the Contributor Coverage Ratio, that is, the ratio of contributors to employees, for companies that are open source-first vs. those that are company-first or started at the same time.

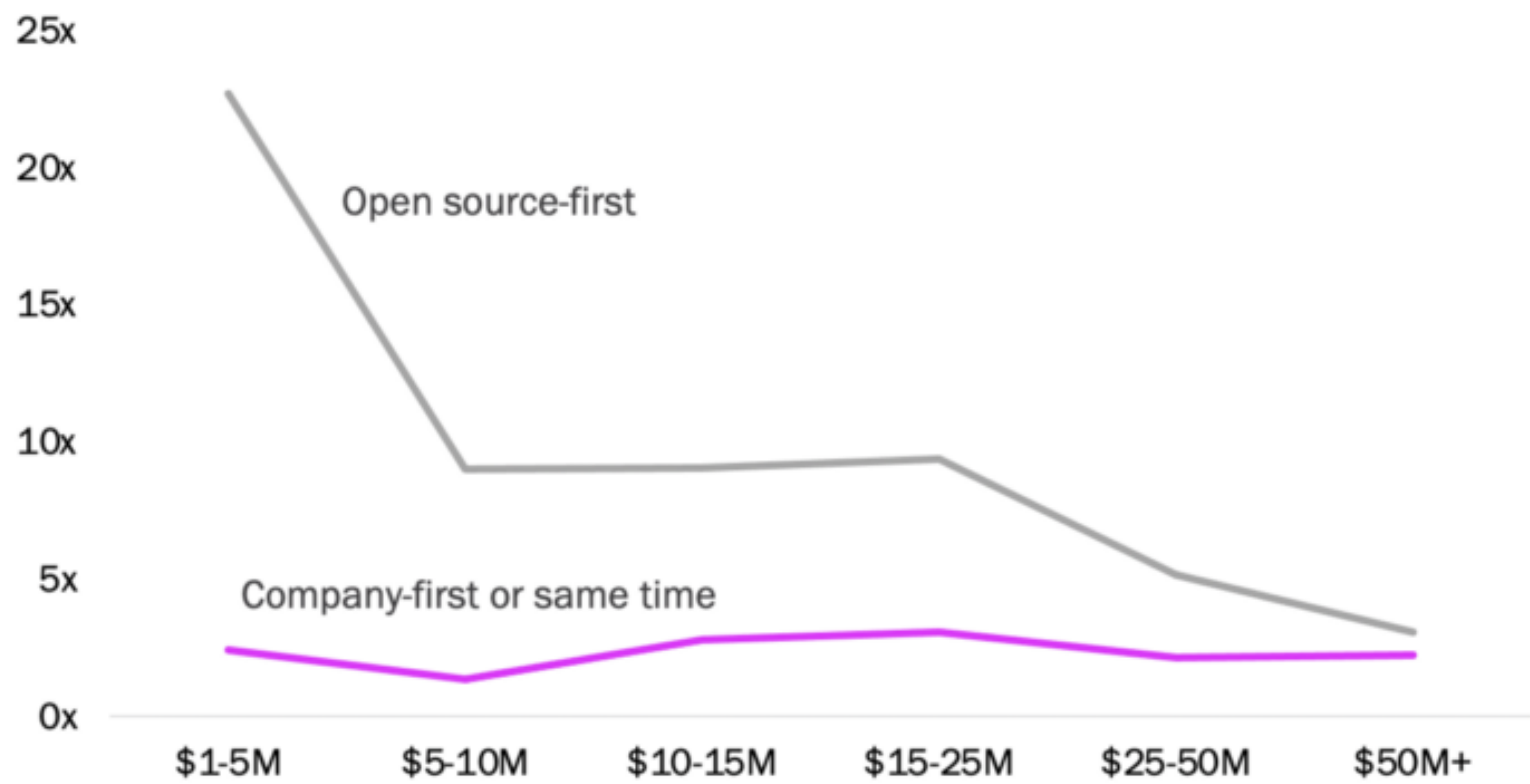
At seed stage, those that are open source-first had a significant advantage where they had a 23x Contributor Coverage Ratio. But by growth stage, this advantage has evaporated as both groups have Contributor Coverage Ratios just over 2–3x.

If building a community of contributors is important to the strategy, then to be successful, companies that leverage open source as a go-to-market strategy need to be able to attract contributors at a rate commensurate with their company growth. Even those that are open source first can't simply focus on commercialization and need to continue to invest in community.



Contributor growth should keep up with the company growth

(Ratio of # contributors to # employees)



Successful open source companies view their community as a source of leads for customers, talent, and support. Rasa recognizes that it's crucial to attract contributors, so they make it easy to contribute and celebrate contributors, leading to an excellent Contributor Coverage Ratio of 4x.

In some cases, we've heard about companies strategically leaving unfinished features in their code so that avid users would be enticed to contribute! On the other hand, if the open source community is not growing, then that would also starve the growth of the commercial side and lead to a painful reckoning.

Three metrics to rule them all

Open source is powerful because fostering a developer community drives both development velocity as well as bottoms-up adoption for enterprise sales, thereby creating a formidable moat and often enabling these companies to scale faster than their closed competitors.

But open source complicates the business model by adding more degrees of freedom as well as inviting more public scrutiny on how you build your company. It is very powerful if you get the magic formula working.

As food for thought, here is a framework for early stage founders building open source companies to think about what GitHub metrics to focus on. This is an alpha version, we are continuing to iterate and would love any feedback.





Basis Set Framework for Open Source Metrics on GitHub

	GOOD	BETTER	BEST
Fastest growing adoption metric	Stars	Forks	Issues
Fastest growing contribution metric	Pull Requests / Commits	Contributors	Serious Contributors (with >2 commits)
Contributor Coverage Ratio (for companies that launch open source at the same time or after company formation)	>2x	>2.5x	>4x

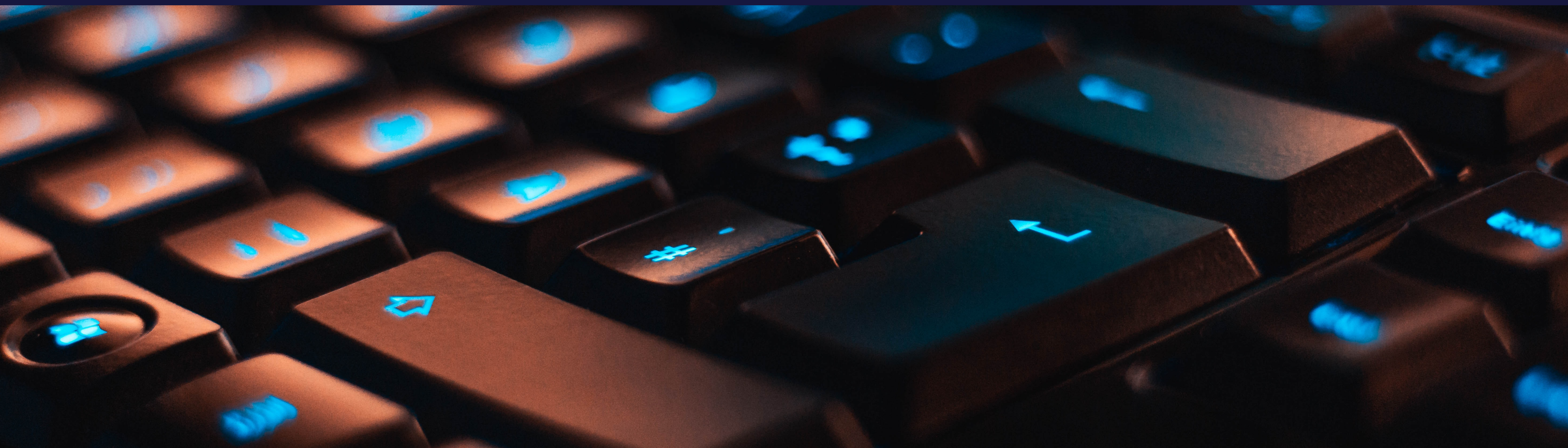
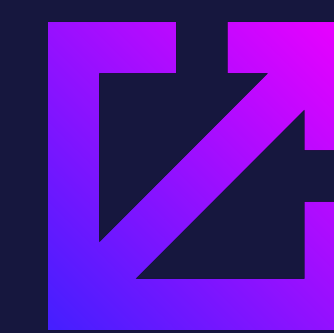


We at Basis Set Ventures are excited about building the next generation of category leaders via open source.

If you have an early stage company or thinking of forming one from your open source project, shoot us a note at chang@basisset.ventures.



Chang Xu
Partner



Methodology

~134 companies founded on or after 2015 that have raised at least \$1M in total funding and are still privately held. See the list of companies included in the Basis Set Open Source Index [here](#). If we've missed your company, let us know [here](#).

...with an open source component to their business. We included both companies for whom open source is core to their strategy (e.g., Fishtown Analytics) as well as those for whom open source is ancillary (e.g., Replicated), but not trivial.

Linking company to GitHub repo:

- As companies frequently have many repos on GitHub, we looked at the repo with the most number of stars and designated it as the primary repo for each company
- We designated GitHub repo by project origin, even if the OSS is not owned by the company, eg Astronomer/Airflow, StreamNative/Pulsar, Altinity/Clickhouse, Ahana/Presto, and Starburst/Presto
- Where applicable, if a company arose out of an open source project and then created a new open source project, we chose to look at the original open source project, eg Observable/D3

Data sources:

- Crunchbase, GitHub, Pitchbook, company websites, LinkedIn, company meetings, manual cleaning

Footnotes

[1] Some 2020 companies are likely still in stealth or haven't yet raised \$1M+, so the count will increase

[2] "Same time" — defined as company formation and open source founding (first commit to GitHub repo) is within 3 months of each other.

[3] We understand that funding is an imperfect metric and there are many nuances, but it is objective and correlates with the viability and success of the business.

[4] Nadia Eghbal calls these projects with high user growth and low contributor growth "Stadiums" in her book "Working in Public: The Making and Maintenance of Open Source Software."

**BASIS SET
VENTURES**



basisset.ventures